LIMNOLOGY
and
OCEANOGRAPHY: METHODS

Limnol. Oceanogr.: Methods 7, 2009, 865–874

Shotgun proteomics: Tools for analysis of marine particulate proteins

Hong-Po Dong[1], Da-Zhi Wang[1], Minhan Dai[1], Leo Lai Chan[1,2], Hua-Sheng Hong[1]*

[1]State Key Laboratory of Marine Environmental Science/Environmental Science Research Center, Xiamen University, Xiamen 361005, China

[2]School of Biological Sciences, The University of Hong Kong, HKSAR, China

Abstract

This study sought a high resolution and high-throughput method to identify and characterize proteins from marine particulate organic matter (POM) using proteomic approaches. The results showed that only a limited number of discrete protein spots were distinguished using two-dimensional electrophoresis (2-DE). Most protein spots were faint and small in 2-DE gels, with a heavy unresolved smeared staining background, indicating 2-DE was not a good high resolution method to separate particulate proteins for identification and characterization. The shotgun proteomic approach combining one-dimensional electrophoresis and capillary liquid chromatography-tandem mass spectrometry as well as the NCBI protein database search was successfully applied to identify and characterize particulate proteins. Using this approach, 737 proteins matching one or more peptides were detected in a POM sample collected from the 41 m water layer in the basin area of the western South China Sea. Of these, 184 were identified as high-confidence proteins matching two or more peptides, including photosynthetic proteins, transporters, molecular chaperones, and porins. In addition to these proteins with known functions, a significant number of novel proteins (accounting for ~30% of the proteins identified) were also detected. The identification of a large number of high-confidence proteins in the POM sample demonstrated that the shotgun proteomic approach is reliable and feasible for the study of particulate proteins and will provide a powerful tool to comprehensively investigate the nature and dynamics of POM in the ocean.

Introduction

Particulate organic matter (POM) is an important carbon pool in the ocean, which not only regulates the air-sea carbon flux but also influences the global carbon cycle. Previous studies demonstrate that proteinaceous materials are a major component of POM (Setchell 1981; Long and Azam 1996). Among others, particulate combined amino acids (PCAAs), a mixture of cellular proteins and detrital combined amino acids, are considered to be one of the major constituents of POM in oceanic surface waters (Tanoue 1992, 1996), and are also the largest component (40% to 50%) of POC in sinking POM (Hedges et al. 2001). PCAAs in POM include at least four chemical forms: proteins, acidic materials containing peptides, protein/peptide-specific dye-stainable low-molecular-mass materials, and non-stainable materials. However, to date, the protein characteristics of POM are not well documented, although a few discrete proteins are separated and identified from POM (Saijo and Tanoue 2004). POM is still a poorly understood organic pool in terms of its source, its chemical composition as well as its transformation from the surface to the ocean depth.

Proteins are organic compounds made of amino acids arranged in a linear chain and folded into a globular form. The composition and sequence of amino acids in proteins can bear information about their source organisms and cellular function including structural organization, stress response, and energy production and conversion. The characterization of the proteins from POM will improve our understanding of the sources and mechanisms that control the cycling and long-term preservation of organic matter (Nunn and Timperman 2007), so a higher resolution and high-throughput analytical

*Corresponding author: E-mail: dzwang@xmu.edu.cn
Tel: + 86 592 2186016, Fax: + 86 592 2180655

Acknowledgments

We thank the captain and crew of the R/V *DongFangHong 2* for their assistance during the cruise. This study was supported by the National Natural Science Foundation of China through grants 40821063, 40376032, and 40476053, and the Ministry of Science and Technology through grant 2008DF100440; D.-Z. Wang was also supported by the Program for New Century Excellent Talents in Xiamen University. Prof. John Hodgkiss is thanked for his assistance with English.

technique for the particulate proteins has become very necessary. Tanoue (1992, 1996) introduced a gel-based separation method, one-dimensional sodium dodecyl sulfate polyacrylamide gel electrophoresis (1D SDS-PAGE) to separate particulate proteins in the water columns of the northern North Pacific and the Equatorial Pacific, and revealed that particulate proteins can be divided into 2 groups according to their electrophoretic patterns, i.e., background proteins and specific proteins. The former group was unable to be resolved using 1D SDS-PAGE and consisted of a large number of proteins of widely ranging molecular mass, each present at a relatively low level. The latter, which have a limited range of molecular masses, are superposed on the background proteins. However, this method could not separate particulate proteins efficiently due to its inadequate resolution of complex marine POMs. Saijo and Tanoue (2004) applied two-dimensional electrophoresis (2-DE), a new method for separating proteins by isoelectric point and subsequently by molecular mass, to investigate particulate proteins and separate 23 discrete proteins from POM. Recently, this method was also used to analyze the metaproteome of marine microbial communities, the major component of POM (Kan et al. 2005). Although these pioneer works provided useful information on the protein composition of POM, however, it was still very difficult to resolve effectively a majority of proteinaceous material by one-dimensional electrophoresis (1-DE) or 2-DE in terms of high complexity and very low abundance of particulate proteins in POM, which hindered further survey of the variability, universality, and identification of proteins. So far, very few proteins have been identified from POM using the gel-based method, and the protein characteristic of marine POM is still an enigma. Apparently, new techniques for separating and identifying the particulate proteins are essential for characterizing chemical and biological features of POM.

"Shotgun proteomics" is a newly developed technique designed to analyze the large scale of proteins in a complex protein mixture, and it has been successfully applied to characterize the protein profiles of certain cells or organisms (Beausoleil et al. 2004; Adachi et al. 2006; Zhang et al. 2007). In contrast to the traditional gel-based method, which separates the protein mixture prior to enzymatic cleavage, this approach produces a mixture of peptides through global digestion of the pool of proteins, and then the mixture of peptides are separated and analyzed usually with multidimensional liquid chromatography (LC) coupled to a tandem electrospray ionization mass spectrometer. Moreover, this method is largely unbiased, allowing both high- and low-abundance proteins, and proteins with extremes in pI and Mr, to be identified with equal sensitivities in the low femtomole level, and also allowed for the analysis of membrane proteins (Wu and MacCoss 2002). Recently, this method has also been applied to study the proteomics of environmental samples (Powell et al. 2005; Ram et al. 2005; Schulze et al. 2005; Sowell et al. 2009) and has displayed its powerful capacity for analyzing proteins in complex samples.

Here, we first applied 2-DE to separate the particulate proteins in a POM sample from the South China Sea (SCS), and we found that only a limited number of protein spots were observed but the heavy background staining and very low concentration of the proteins in the gel interfered with mass spectrometric identification. So, a new strategy, "shotgun" mass spectrometry (MS)–based proteomics was applied to characterize proteins in marine POM. To our knowledge, this study is the first attempt to apply this approach to characterize proteins of POM in oceanic waters. Using this approach, 737 proteins were identified and characterized in a POM sample from the 41 m water layer in the SCS, and their subcellular localization and biological processes were interpreted. This study provides a powerful tool for future comprehensive investigations of the protein composition of marine POM.

Materials and procedures

Sample collection—POM samples from different water layers were collected at Station TS1 (14°15.495′N, 111°45.563_E; water depth, 2390 m) during a summer cruise to the western SCS from 14 August to 14 September 2007. The surface POM was filtered onto 150 mm GF/F glass filter using a pumping system. Samples from other depths were collected onto 150 mm GF/F glass fiber filters (Whatman) using an in situ large volume water transfer system sampler (McLane), the filtered seawater volume at each depth being surface (5m), 260 L; 15 m, 800 L; 41m, 1000 L; 100 m, 860 L. Any visible zooplankton were removed with clean forceps, and all samples were kept frozen at –80°C aboard ship.

Protein extraction and quantification in POM—To release specific proteins bound within or trapped in refractory organic or inorganic matrices, three procedures were used to extract particulate proteins: (1) a GF/F filter with POM was thawed and cut into chips. The chips were suspended in 5 mL lysis buffer consisting of urea (7 mol L^{-1}), thiourea (2 mol L^{-1}), CHAPS (2% w/v), Triton-X100 (2% v/v), carrier ampholytes (2% v/v, pH3-10), dithiothreitol (DTT 1% w/v), and protease inhibitor cocktail (1% v/v), and shaken in an ice-cold water bath for 30 min using an ultrasonic shaker. The suspension was placed on ice and lysed with the help of a sonicator using a micro-probe at high power with 6-s pulses for 8 min; (2) the GF/F filter chips were mixed with 5 mL lysis buffer and homogenized into a slurry with a mortar and pestle; and (3) the GF/F filter chips were put into a 2 mL snap-top microtube containing 1 mL lysis buffer and glass beads, and then homogenized by violently agitating the microtube using a Mini-BeadBeater. All slurries were incubated at 25°C for 1 h and then centrifuged at 20,000*g* for 30 min at 10°C. The supernatant was precipitated with ice-cold 20% trichloroacetic acid in acetone (TCA/acetone, 1:5 v/v) for at least 12 h at –20°C. The mixture was centrifuged at 20,000*g* for 30 min at 4°C, and then the pellets were rinsed twice with ice-cold acetone, and air-dried. Finally the powder was dissolved in 70 µL rehydration buffer containing 7 mol L^{-1}urea, 2 mol L^{-1} thiourea, and CHAPS (4% w/v).

The protein contents in the POM concentrates were quantified using the Bradford method and bovine serum album as the protein standard. The concentrate was stored in a freezer (–80°C) until analysis. Unless otherwise noted, all the reagents used in this study were electrophoretic grade.

Two-dimensional SDS-PAGE of particulate proteins—For each POM sample, 100 µg protein was mixed with a rehydration buffer and then loaded onto IPG strips of linear pH gradient 4-7. Rehydration and subsequent isoelectric focusing were conducted using the Ettan IPGphor III Isoelectric Focusing System (Amersham Biosciences). Isoelectric focusing was performed in the following manner: 13 h at 50 V, 2 h at 100V, 2 h at 200V, 1 h at 500V, 2 h at 1000V, 2 h at 4000V, and 6 h at 8000V. After the first dimension run, each strip was equilibrated with about 10 mL of a solution containing 50 mmol L^{-1} Tris at pH 8.8, 6 mol L^{-1} urea, 30% glycerol, 2% SDS, 1% DTT, and a trace amount of bromophenol blue for 20 min. The strip was then placed in fresh equilibration buffer containing 2.5% iodoacetamide (instead of DTT) for another 20 min. Subsequently a 12.5% SDS-PAGE second dimension was performed. Silver staining was performed following the method of Wang et al. (2009). Images were made using a gel-documentation system on a GS- 670 Imaging Densitometer from Bio-Rad (USA), and 2-DE electrophoretogram matching software.

One-dimensional SDS-PAGE and in-gel digestion of particulate protein—The particulate protein was loaded onto a 5% to 12% Bis-Tris gel according to the method of Laemmli (1970). Electrophoresis was carried out using the Tris-glycine–SDS buffer system (25 mmol L^{-1} Tris, 192 mmol L^{-1} glycine, and 0.1% SDS) on a Hoefer SE 600 apparatus (Amersham) at 160 v /gel, until the dye front reached the bottom edge of the gel. Low molecular weight standards were used in the gel, and it was fixed and stained with colloidal Coomassie Brilliant Blue G-250 (CBB, Bio-Rad). The gel lane was cut into four equal pieces and subjected to in-gel tryptic digestion as described by Wilm et al. (1996). Briefly, the gel pieces were destained and washed, and then, after dithiothreitol reduction and iodoacetamide alkylation, the proteins were digested with trypsin (modified sequencing grade; Promega) overnight at 37°C. The resulting tryptic peptides were extracted from the gel pieces with 60% acetonitrile, 0.1% trifluoroacetic acid, and 100% acetonitrile. The extracts were dried completely using a SpeedVac and stored in a freezer (–80°C) until liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis.

Mass spectrometry analysis—The dried peptide samples were redissolved in 20 µL 0.1% formic acid and injected onto a Zorbax 300SB C18 peptide trap (Agilent Technologies), and desalted with 0.2% formic acid at a flow rate of 10 µL min^{-1} for 20 min. Peptides were eluted from the trap and separated on a reversed phase C18 column (0.15 mm × 150 mm, Column Technology) with a linear gradient of 0% to 50% mobile phase B (0.1% formic acid-84% acetonitrile) in mobile phase A (0.1% formic acid) over 60 min at 65 µL min^{-1}. LC-MS/MS measurements were made with a linear trap quadrupole (LTQ) mass spectrometer (Thermo Finnigan) equipped with a microspray source. The LTQ mass spectrometer operated in the data-dependent mode with the following parameters: spray voltage (3.4 kV), spray temperature 160°C, full scan m/z range (400-1800). To prevent repetitive analysis of the same abundant peptides, dynamic exclusion was enabled with a repeat count of 2 and an exclusion duration of 1.5 min on the LTQ. The LC-MS system was fully automated and under the direct control of an Xcalibur software system (Thermo Finnigan). The ten most intense ions in every full scan were automatically selected for MS/MS.

Bioinformatics analysis of particulate proteins—In this study, a large protein database of potential groups of organisms at TS1 was created based on the protein bank of the National Center for Biotechnology Information (NCBI) Web site according to the in situ planktonic community data and the previous research concerning the composition of detrital organic matter, plus consideration of experimental conditions (Yamada and Tanoue 2006; Kaiser and Benner 2008). Totally, 780573 protein entries from the majority of marine planktonic communities including marine bacteria and archaea, marine phytoplankton, and zooplankton as well as those closely related environmental samples with marine bacteria or archaea were selected to establish the protein database (Table 1). It should be borne in mind that the database created covered only that portion of the organisms from which particulate proteins originated, due to the limited information of marine organisms in the gene bank.

The MS/MS spectra from the LTQ dataset were searched against the above mentioned database using the SEQUEST algorithm. All SEQUEST searches were performed using Bioworks 3.2 software (Thermo Finnigan) with the following parameters: fully tryptic peptide, Parent Mass Tolerance, 1.4; and Peptide Mass Tolerance, 1.5. The protein identification criteria that we used were based on Delta CN (≥0.1) and Xcorr (one charge ≥ 1.9, two charges ≥ 2.2, three charges ≥ 3.75). In addition, since false positive identifications mainly occur within one peptide matched proteins, the identified proteins matched by two or more peptides are considered as high-confidence detection (He et al. 2005; Ram et al. 2005).

Protein identifications were extracted from the SEQUEST.out file using the in-house software BuildSummary, which combined the peptide sequences into proteins and deleted redundant proteins as described by Wu and Yates (2003). Their molecular functions and biological processes were categorized using the Blast2GO tool on www.geneontology.org (Conesa et al. 2005). Homology search was performed at nonredundant protein database downloaded from NCBI website. The Blast2GO extracts the GO terms associated with homologies and returns a list of GO annotations represented as hierarchical categories of increasing specificity. The data presented in Fig. 1 represented the level 3 analysis for biological processes. It is noted that some proteins might be categorized into different biological processes. Protein subcellular location was predicted

Table1. Protein databases of the potential plankton communities in Station TS1 downloaded from the NCBI Web site.

Protein database	Number of protein entries	Protein database	Number of protein entries
Pseudomonas	164718	environmental samples of Euryarchaeota	8938
Flavobacteria	101876	marine archaeal group 1	6058
Erythrobacter	18507	Nanoarchaeota	1166
Photobacterium	34468	Cyanobacteria	343271
Coxiella	22714	Cryptophyta	3323
Environmental samples of Alphaproteobacteria	314	Chlorophyta	25357
Environmental samples of Betaproteobacteria	985	Haptophyceae	862
Environmental samples of Gammaproteobacteria	532	Dinophyceae	2422
Environmental samples of Archaea	4538	stramenopiles	12662
Environmental samples of Crenarchaeota	3737	Crustacea	24125

using the Proteome Analyst - Subcell Specialization Server 2.5 (http://pasub.cs.ualberta.ca:8080/pa/Subcellular) (Lu et al. 2004).

Assessment

Particulate protein extraction and two-dimensional electrophoresis—Three different procedures were conducted to extract particulate proteins from a surface POM sample at Station TS1. Procedure 1 (P1) provided the highest concentration of protein (average protein content: 6.07 μg L^{-1}) from the surface POM sample. When comparing the 2-DE electrophoretograms obtained using the three extraction procedures, P1 exhibited the best result with relatively high resolution and intensity of protein spots, and also a high number of protein spots (data not shown), indicating that P1 released more particulate proteins from refractory organic or inorganic matrices compared with other two procedures. So P1 was used to extract particulate proteins from all POM samples in this study. 2-DE of POM samples from the 15, 41, and 100 m water layers was carried out and the results are shown in Fig. 2. More than 100 discrete proteins of M_r 14-96 kDa were observed in the acidic region (pI range 4-6) (Fig. 2). However, the electrophoretograms of each POM sample presented an unresolved smeared staining background in the acidic region and a very limited number of discrete protein spots were distinguished from the POM samples (Fig. 2). Moreover, the majority of protein spots were faint and small in the 2-DE gels. Saijo and Tanoue (2004) applied 2-DE to investigate particulate proteins from Pacific surface waters, and they found no spots in 2-DE gels visualized by the silver staining method due to the smeared and heavy background staining. Using the CBB-R staining method, they separated a limited number of discrete protein spots (23) from 19 surface POM samples and identified one protein using N-terminal amino acid sequence analysis. However, the 2-DE method failed to resolve acidic and low molecular mass materials from the POM. These results indicated that the 2-DE method was inadequate to separate particulate proteins, especially those low abundance proteins.

Identification and characterization of particulate proteins using shotgun proteomics—Particulate proteins in a POM sample from the 41 m water depth at Station TS1 was prepared using P1 and separated by 1D SDS-PAGE (Fig. 3). The electrophoresis visualized using CBB stain showed a smeared staining pattern throughout the molecular mass range, and no clear band was recognized due to the heavy background staining. For protein identification, four equal slices were cut from the 1D SDS-PAGE gel, and each slice was analyzed using the LTQ mass spectrometer. Peak list files obtained from the four fractions were processed separately, and the peptide sequences were identified. A total of 737 proteins matching one or more peptides were identified from the 41 m POM sample after removal of contaminants (keratins, trypsin) and redundant proteins. For confident detection, 184 particulate proteins, which matched two or more peptides, were identified, accounting for about 25% of identified proteins in the POM sample from the 41 m water layer. About 30% of the proteins identified were novel proteins that are the products of genes previously annotated as "hypothetical." Based on the BLASTP algorithm, these proteins lacked significant homology to proteins with functional assignments. The NCBI accession number, protein name, number of peptides used in the identification, sequence coverage (%), theoretical MW and pI, and the subcellular location of particulate proteins are listed in Appendix 1.

The theoretical pI and MW values of the identified proteins were calculated using DTASelect software. About 70% of the proteins identified had theoretical pI values in the pH 4-6.9 range, with a theoretical MW between 10 to 60 kDa. However, a few proteins with extreme values of pI or MW were also present in the POM sample (Fig. 4). For example, 62 proteins with theoretical MW > 100 kDa and 23 small proteins with theoretical MW < 10 kDa were identified. While 14 acidic proteins with a theoretical pI of <4.5, and 28 basic proteins with a theoretical pI of >10.0 were identified. These proteins represent a group of proteins that are rarely identified using 2-DE. Large numbers of acidic proteins with a theoretical pI of 4-6.9 were detected, suggesting that in this POM sample, acidic proteins or proteinaceous materials

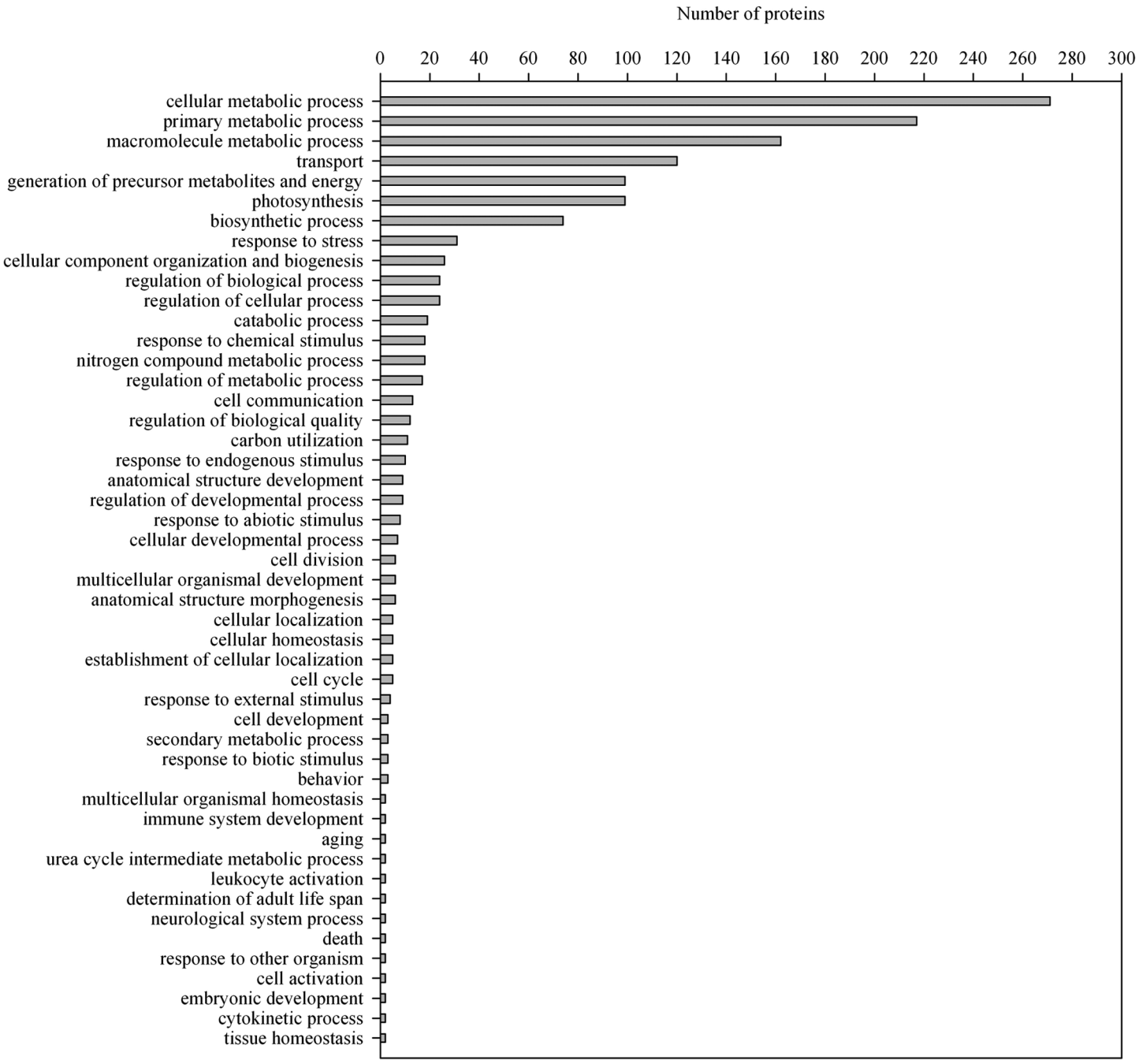

Fig. 1. Gene ontology (GO) biological process terms for the proteins identified in a POM sample from the 41 m water depth in the South China Sea. GO is a structured, controlled vocabulary that describes gene products in terms of their associated biological processes, cellular components, and molecular functions in a species-independent manner.

dominated the bulk particulate proteins. The present result was consistent with the characteristics of electrophoretograms of POM samples from the 15, 41, and 100 m water depths, that unresolved smeared staining background and a limited number of discrete protein spots mainly appeared on the acidic region (Fig. 2). Saijo and Tanoue (2004) proposed that acidic proteins with negative charge in seawater may facilitate association between particulate proteins with other organic or inorganic materials.

In this study, the proteins identified were divided into 2 groups according to their biological origins: prokaryotic proteins and eukaryotic proteins, and subsequently classified into subcellular compartments using Proteome Analyst - Subcell Specialization Server 2.5. Of the prokaryotic particulate

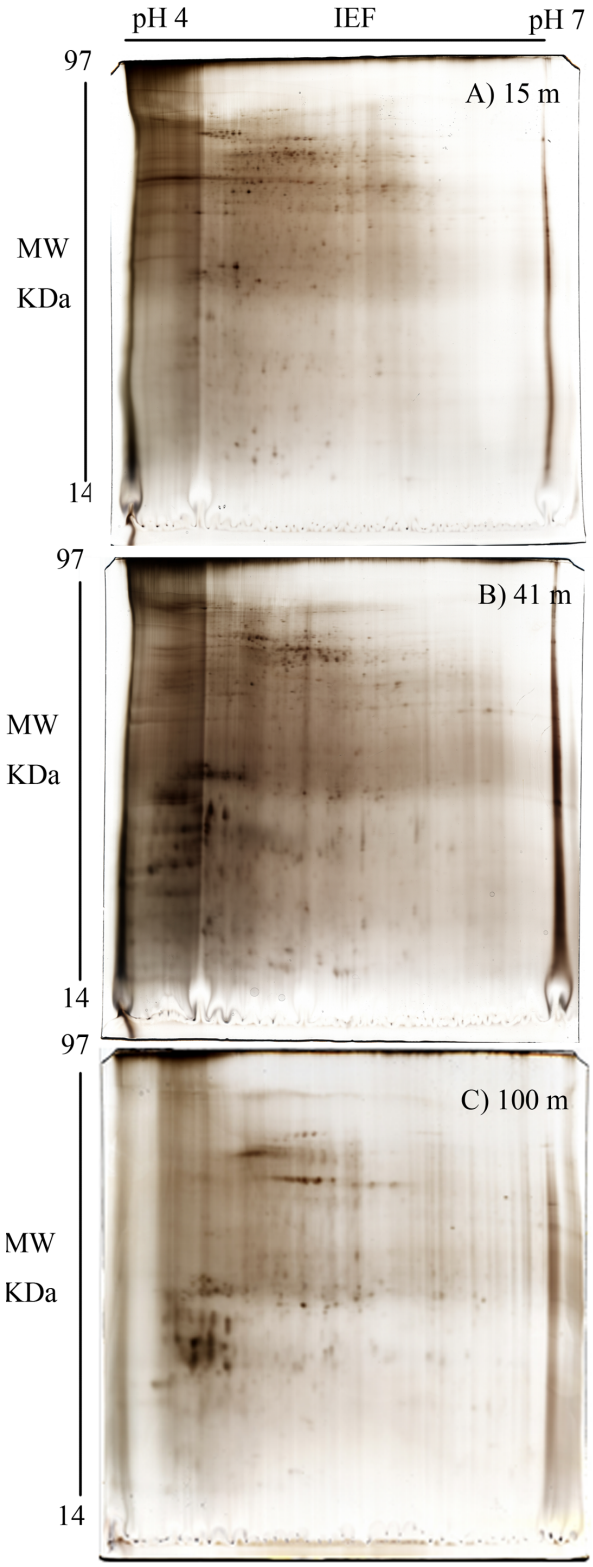

Fig. 2. Two-dimensional electrophoretograms of particulate proteins from the 15 m (A), 41 m (B), and 100 m (C) water layers in the South China Sea. Proteins on the gels were stained using silver staining. About a 100 µg bovine serum albumin equivalent amount of electrophoretic sample was loaded for each gel.

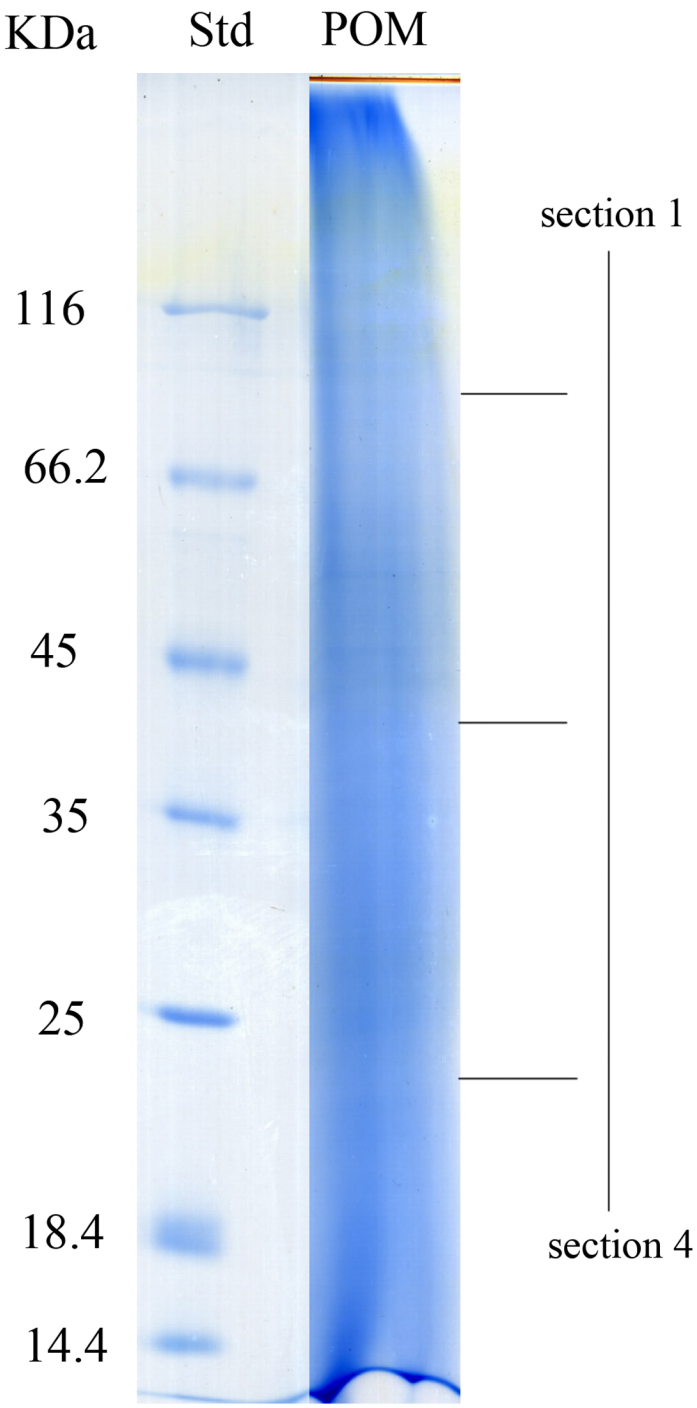

Fig. 3. One-dimensional SDS-PAGE electrophoretogram of particulate proteins in a POM sample from the 41 m water layer in the South China Sea. The gel was stained with colloidal coomassie blue and cut into four pieces. About a 60 µg bovine serum albumin equivalent amount of electrophoretic sample was loaded.

proteins, those with unknown cellular localization accounted for about 41% of the prokaryotic proteins in the POM sample (Fig. 5A). About 26% of prokaryotic proteins were membrane proteins including proteins associated with inner and outer

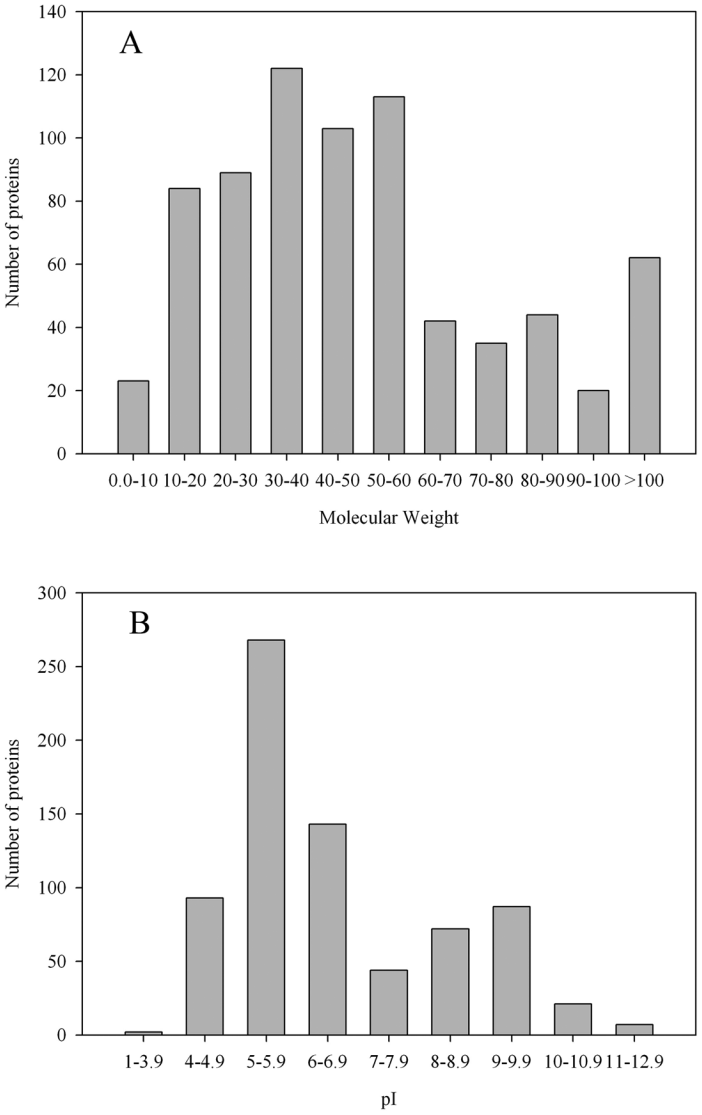

Fig. 4. Distribution of molecular weights and isoelectric points of the proteins identified in a POM sample from the 41 m water layer in the South China Sea. (A) Molecular weight distribution and (B) pI distribution.

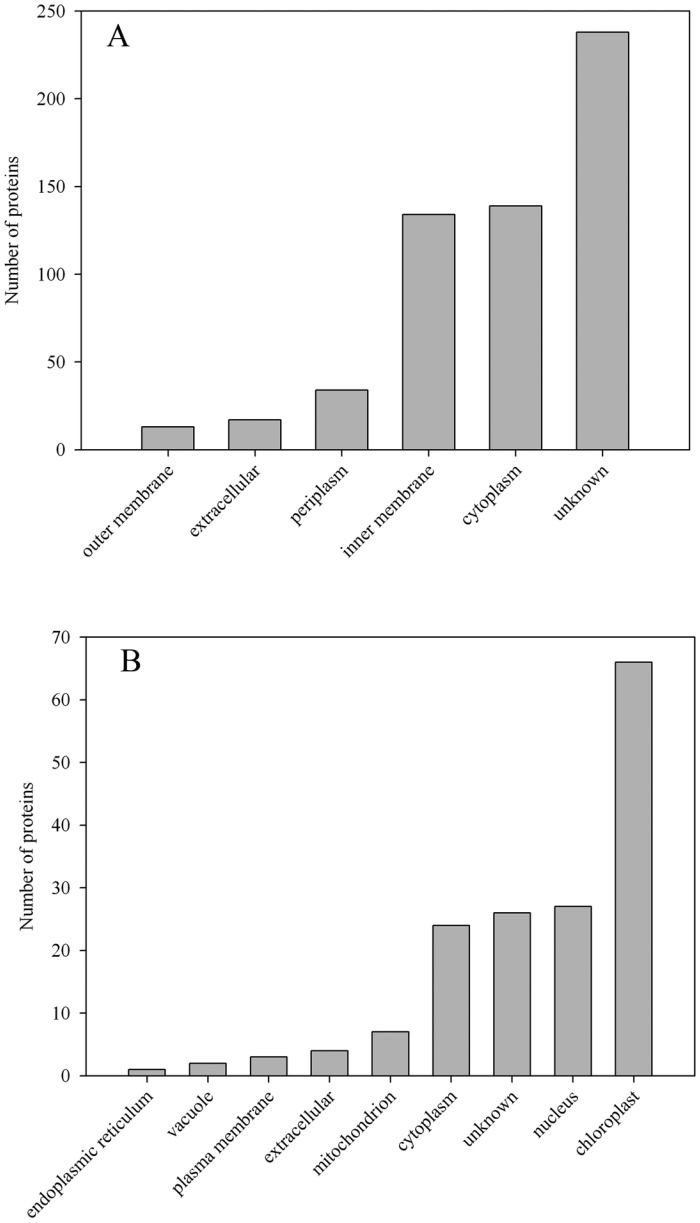

Fig. 5. Classification of the proteins identified by subcellular localization in a POM sample from the 41 m water depth in the South China Sea. (A) Subcellular localization for proteins from prokaryotes and (B) subcellular localization for proteins from eukaryotes.

membranes, such as ATP synthases, light-harvesting proteins, photosystem II PsbC proteins, cytochromes, porins, TonB-dependent siderophore receptors, and transporters. Cytoplasmic proteins accounted for 24% of the prokaryotic proteins in the POM sample. Seventeen proteins from prokaryotes were classified as extracellular proteins, including proteinases, peptidases, glucanases, galactosidases, and hydrolases. A total of 34 proteins from prokaryotes were classified as periplasmic proteins, including transporters, dehydrogenases, periplasmic substrate binding proteins, glutamyltranspeptidase, murein transglycosylase, copper-binding protein and iron(III) dicitrate-binding periplasmic protein. Chloroplastic proteins comprised 41% of the eukaryotic particulate proteins in the POM sample while cytoplasmic and nuclear proteins accounted for 17% of the eukaryotic particulate proteins (Fig. 5B). Detection of proteins from different subcellular components attested to the validity of the unbiased, comprehensive shotgun proteomic approach.

The particulate proteins identified were functionally categorized based on universal Gene Ontology (GO) annotation terms using the Blast2GO tool against the non-redundant protein database. As shown in Fig. 1, the proteins identified were involved in 48 biological processes, including transport, generation of precursor metabolites and energy, photosynthesis,

cellular component organization and biogenesis, nitrogen compound metabolic process, carbon use, cell communication, and other important processes. For example, 120 proteins that were involved in transport (such as the urea ABC transporter, phosphate transporter, sulfate transporter and ammonium transporter) were identified, of which 14 were high-confidence proteins, and played critical roles in the transport of various nutrients; 18 proteins involved in nitrogen compound metabolism were detected, including methylenetetrahydrofolate dehydrogenase, acetolactate synthase, urease, glutamine synthetase, and aspartate carbamoyltransferase; 11 proteins involved in carbon use were detected, including periplasmic β-glucosidase, 5-carboxy-2-hydroxymuconate semialdehyde dehydrogenase, ribulose 1,5-bisphosphate carboxylase/oxygenase and phosphoribulokinase; and 99 proteins were identified as being involved in photosynthesis and the generation of precursor metabolites and energy (Fig. 1), including ATP synthase, light-harvesting complex protein, cytochrome, and phycoerythrin. These results suggest that high biological activities occurred in the 41 m layer.

Discussion

Proteomic profiling either of protein mixtures fractionated using 1-DE, or of unfractionated protein mixtures using protease digestion and subsequent LC-MS/MS analysis, is referred to as 'shotgun' proteomics. Compared to the 2-DE based MS/MS approach, the gel-based LC-MS/MS approach does not require the isoelectric focusing step and is more effective in the study of hydrophobic proteins, such as membrane proteins (Wu and Yates 2003). In the past few years, this technique has been comprehensively applied in studies of the proteomics of various samples. Adachi et al. (2006) identified 1543 proteins in human urine using the 1-DE based LC-MS/MS approach. Zhang et al. (2007) characterized 458 GO annotated membrane proteins with extremely high confidence from mice macrophage using the same approach. In addition, this approach is also used for the analysis of phosphopeptides from nuclear Hela cell lysates (Beausoleil et al. 2004), and mice embryo forebrains and midbrains (Ballif et al. 2004). However, only a few studies have analyzed environmental samples using shotgun proteomics (Powell et al. 2005; Ram et al. 2005; Schulze et al. 2005). Powell et al. (2005) analyzed massive numbers of CID spectra of the seawater protein trypsin digests produced by HPLC-MS/MS analysis using two different methods: one was to search the CID spectra against a nonredundant DNA and protein database using the software algorithm SEQUEST; the other was to analyze these proteins with the same peptidic CID spectra using de novo sequencing analysis to determine amino acid sequence information. However, either of them was not an ideal method for large-scale protein identification due to the lack of sequences of dissolved proteins in database or the requirement of high quality CID spectra and laboriousness. Recently, Ram et al. (2005) detected 5994 proteins from the five most abundant species in a natural acid mine drainage microbial biofilm community using the MS-based proteomic method. However, after combining community genomic analysis, they finally identified 2033 proteins matching two or more peptides as high-confidence proteins. Using the 'shotgun' MS-based proteomic approach and a variety of detection criteria and protein databases from organisms not present in the biofilm community, they estimate the likelihood of false positive protein identification and find that the false-positive rate at the one peptide level is 16%, but this decreases to 6.2% when considering proteins matched by at least two peptides.

In the present study, the conventional 1-DE combined with the LC-MS/MS approach was used to separate and identify particulate proteins from the SCS, and 737 proteins matching one or more peptides were identified from the POM sample, with 184 of them being identified as high-confidence proteins matching two or more peptides. It should be emphasized that the protein database of marine organisms is seriously limited since the genomes of only a limited number of marine organisms have been sequenced (Powell et al. 2005), and also database matching software is intrinsically limited because only those proteins present in the database can be matched. Thus, the particulate proteins identified in this study were only a part of the bulk particulate proteins in POM. Although these proteins were screened with stringent search parameters (He et al. 2005), because of the high diversity and uncertainty of source organisms of POM, it was not possible to estimate the false-positive rate of protein identification, and so there might be a high false-positive rate of identification within these one peptide-matched proteins. It is also possible that a large number of proteins (~75%) identified with one peptide were derived from low-abundance proteins or small peptides due to the extensive biodegradation during the sinking of POM (Hedges et al. 2001; Saijo and Tanoue 2005).

This study identified 80 photosynthesis proteins from POM at 41 m layer using the shotgun proteomic approach, and majority of them were integral-membrane protein complexes from the photosystem I and II reaction center. Moreover, 25 ATP synthases localized in the thylakoid or photosynthetic membranes were also identified from POM. These results suggested that high photosynthetic activities occurred in the 41 m layer. It is noteworthy that this 41 m layer was the maximum Chlorophyll *a* concentration layer during the sampling period, and was characterized by maximum chlorophyll concentration and phytoplankton biomass. Transporters are another dominant component of the particulate proteins identified (Appendix 1), and the transporters involved in cyanobacterial urea uptake were frequently detected, suggesting urea is an important nitrogen source for cyanobacteria in the oligotrophic SCS. Transporters for ammonium in cyanobacteria were also identified but nitrate and nitrite transporters were not detected. The results supported the previous view that cyanobacteria prefer to use

reduced forms of nitrogen such as urea, ammonium, and amino acids than oxidized sources such as nitrate and nitrite, forcing other bacteria to use inorganic nitrogen at a higher energetic cost in the oligotrophic environment (Zubkov et al. 2003; Garcia-Fernandez et al. 2004). Transporters involved in the uptake of phosphate were also frequently detected. It is known that these transporters are induced in bacteria and phytoplankton during phosphate stress conditions (Dyhrman and Haley 2006; Martiny et al. 2006). Our results suggested that organisms in the POM were stressed by phosphorus, which was consistent with the phosphorus status at the surface in the SCS. Concentration of phosphorus is generally ~5 nM during summer in the SCS (Ning et al. 2004). Polyubiquitin and ubiquitin proteins were among the most prevalent proteins identified in this study, and the chaperone proteins hsp70, groEL, dnaK were also detected. The identification of these ubiquitin and chaperone proteins suggested that protein refolding, stability, and degradation may be necessary for various organisms in the maximum Chl *a* concentration layer in the SCS as a result of exposure to environmental stresses. Based on the above discussion, the metaproteomic analysis of marine POM has provided new biological information and insights into environmental stress in the ocean.

Overall, the identification of a large number of high-confidence proteins involved in photosynthesis in the POM in the euphotic zone demonstrated that the "shotgun proteomic" approach is reliable and feasible for the study of particulate proteins, and provides a potentially powerful tool to comprehensively investigate the protein composition of POM. In future, more effort should be devoted to the low-abundance proteins or small peptides, to obtain more useful information for the elucidation of the source of POM and the mechanisms involved in the recycling of POM in marine systems.

References

Adachi, J., C. Kumar, Y. L. Zhang, J. V. Olsen, and M. Mann. 2006. The human urinary proteome contains more than 1500 proteins, including a large proportion of membrane proteins. Genome Biol. 7:R80.1-R80.16.

Ballif, B. A., J. Villen, S. A. Beausoleil, D. Schwartz, and S. P. Gygi. 2004. Phosphoproteomic analysis of the developing mouse brain. Mol. Cell Proteom. 3:1093-1101.

Beausoleil, S. A., and others. 2004. Large-scale characterization of HeLa cell nuclear phosphoproteins. Proc. Natl. Acad. Sci. U.S.A. 101:12130-12135.

Conesa, A., S. Gotz, J. M. Garcia-Gomez, J. Terol, M. Talon, and M. Robles. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674-3676.

Dyhrman, S. T., and S. T. Haley. 2006. Phosphorus scavenging in the unicellular marine diazotroph *Crocosphaera watsonii*. Appl. Environ. Microbiol. 72:1452-1458.

Garcia-Fernandez, J. M., N. T.Marsac, and J. Diez. 2004. Streamlined regulation and gene loss as adaptive mechanisms in prochlorococcus for optimized nitrogen utilization in oligotrophic environments. Microbiol. Mol. Biol. R. 68:630-638.

Hedges, J. I., J. A. Baldock, Y. Gelinas, C. Lee, M. Peterson, and S. G. Wakeham. 2001. Evidence for non-selective preservation of organic matter in sinking marine particles. Nature 409:801-804.

He, P., and others. 2005. The human plasma proteome: Analysis of Chinese serum using shotgun strategy. Proteomics 5:3442-3453.

Kaiser, K., and R. Benner. 2008. Major bacterial contribution to the ocean reservoir of detrital organic carbon and nitrogen. Limnol. Oceanogr. 53:99-112.

Kan, J., T. E. Hanson, J.M. Ginter, K. Wang, and F. Chen. 2005. Metaproteomic analysis of Chesapeake Bay microbial communities. Saline Syst. 1 [doi: 10.1186/1746-1448-1-7].

Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680-685.

Long, R. A., and F. Azam. 1996. Abundant protein-containing particles in the sea. Aquat. Microb. Ecol. 10:213-221.

Lu, Z., and others. 2004. Predicting subcellular localization of proteins using machine-learned classifiers. Bioinformatics 20:547-556.

Martiny, A. C., M. L. Coleman, M. L.Coleman, and S. W. Chisholm. 2006. Phosphate acquisition genes in Prochlorococcus ecotypes: evidence for genome-wide adaptation. Pro. Natl. Acad. Sci. U.S.A.103:12552-12557.

Ning, X., F. Chai, H. Xue, Y.Cai, C. Liu, and J. Shi. 2004. Physical-biological oceanographic coupling influencing phytoplankton and primary production in the South China Sea. J. Geophys. Res. 109:1-20.

Nunn, B. L., and A. T. Timperman. 2007. Marine proteomics. Mar. Ecol. Prog. Ser. 332:281-289.

Powell, M. J., J. N. Sutton, C. E. Del Castillo, and A. I. Timperman. 2005. Marine proteomics: generation of sequence tags for dissolved proteins in seawater using tandem mass spectrometry. Mar. Chem. 95:183-198.

Ram, R. J., and others. 2005. Community proteomics of a natural microbial biofilm. Science 308:1915-1920.

Saijo, S., and E. Tanoue. 2004. Characterization of particulate proteins in Pacific surface waters. Limnol. Oceanogr. 49:953-963.

———, and ———. 2005. Chemical forms and dynamics of amino acid-containing particulate organic matter in Pacific surface waters. Deep-Sea Res. I 52:1865-1884.

Schulze, W. X., G. Gleixner, K. Kaiser, G. Guggenberger, M. Mann, and E. D. Schulze. 2005. A proteomic fingerprint of dissolved organic carbon and of soil particles. Oecologia 142:335-343.

Setchell, F. W. 1981. Particulate protein measurement in oceanographic samples by dye binding. Mar. Chem. 10:301-313.

Sowell, S. M., and others. 2009. Transport functions dominate

the SAR11 metaproteome at low-nutrient extremes in the Sargasso Sea. ISME J. 3:93-105.

Tanoue, E. 1992. Occurrence and characterization of particulate proteins in the Pacific Ocean. Deep-Sea Res. A 39:743-761.

———. 1996. Characterization of the particulate protein in Pacific surface waters. J. Mar. Res. 54:967-990.

Wang, D.-Z., L. Lin., L. L. Chan., and H.-S. Hong. 2009. Comparative studies of four protein preparation methods for proteomic study of the dinoflagellate *Alexandrium* sp. using two-dimensional electrophoresis. Harmful Algae 8:685-691.

Wilm, M., A. Shevchenko, T. Houthaeve, S. Breit, L. Schweigerer, T. Fotsis, and M. Mann. 1996. Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. Nature 379:466-469.

Wu, C. C., and M. J. MacCoss. 2002. Shotgun proteomics: Tools for the analysis of complex biological systems. Curr. Opin. Mol. Ther. 4:242-250.

———, and J. R. Yates. 2003. The application of mass spectrometry to membrane proteomics. Nat. Biotechnol. 21:262-267.

Yamada, N., and E. Tanoue. 2006. The inventory and chemical characterization of dissolved proteins in oceanic waters. Prog. Oceanogr. 69:1-18.

Zhang, L. B., Y. Lun, D. M. Yan, L. Y. Yu, W. Ma, B. R. Du, and X. Zhu. 2007. Proteomic analysis of macrophages: A new way to identify novel cell-surface antigens. J. Immunol. Methods 321:80-85.

Zubkov, M. V., B. M. Fuchs, G. A. Tarran, P. H. Burkill, and R. Amann. 2003. High rate of uptake of organic nitrogen compounds by prochlorococcus cyanobacteria as a key to their dominance in oligotrophic oceanic waters. Appl. Environ. Microb. 69:1299-1304.

Submitted 16 July 2009
Revised 8 October 2009
Accepted 9 November 2009